PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association Between Dairy Intake And Fracture In An Australian Based Cohort Of Women: A Prospective Study
AUTHORS	Aslam, Hajara; Holloway, Kara; Mohebbi, Mohammadreza; Jacka, Felice; Pasco, Julie

VERSION 1 – REVIEW

REVIEWER	Liisa Byberg Department of Surgical Sciences, Orthopaedics, Uppsala University, Sweden
REVIEW RETURNED	01-Jun-2019

GENERAL COMMENTS	In the submitted manuscript, Aslam and colleagues examine the association of milk intake and total dairy intake with the risk of fractures. A population of 833 women aged 50 years and older were followed for 20-25 years and over this period, 206 major osteoporotic fractures were observed. The main limitation of the study and its presentation, in addition to its limited size, is that it is not clear how incident fractures were ascertained (see also comment below) and how large loss to follow-up there was in the study. Additionally, the associations of milk and total dairy intake with an inflammation marker and with bone turnover markers were examined. The rationale for these analyses is however not explained in the introduction and they are not stated as an aim (this should of course be included). Please also specify the setting for these associations (cross-sectional?). I have the following questions and comments to the authors: 1. Although osteoporosis is associated with a marked increased risk of fracture, most fractures are still preceded by a fall. The largest number of fractures also occur among those who do not have osteoporosis. The introduction should reflect this as well as the bone aspects of fracture risk. 2. The meta analysis cited in ref 18 is rather old and could be excluded. More importantly, results from meta analyses on dietary studies are complicated because of the different dietary context and the different range of intakes (see for instance Barnard ND, Willett WC, Ding EL. The Misuse of Meta-analysis in Nutrition Research. JAMA 2017;318:1435-1436. https://doi.org/10.1001/jama.2017.12083). Reference 19 is problematic due to the large heterogeneity between the

	studies in the main analysis (largest intake group vs the smallest intake group) and the dose-response analysis shows a different result than what is referred to in the introduction. The heterogeneity between studies is due to the differences in the underlying studies with regards to range of milk intake, number of fracture cases, what confounders were adjusted for, and how fractures were ascertained. I encourage the authors to highlight the problems with meta analyses in this context. 3. The interpretation of the results seems to rely heavily on statistical significance. In the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (by ICMJE) and in the STROBE guidelines it is said that authors should avoid relying solely on statistical hypothesis testing, such as *P* values, which fail to convey important information about effect size. I think that considering the small number of fractures in the present study it would be more fair to describe your association between milk intake (and total dairy intake) and fracture risk as u- (or perhaps J-) shaped with higher risks in the no consumption and the high consumption groups and the lowest risk in the low-intermediate consumption groups, however certainty is limited because of the small numer of cases resulting in low precision in the estimates. Further, P-values in table 1 is discouraged (as per STROBE guidelines). 4. How did you select which confounders to adjust for? There seems to be an association between milk intake and educational level – this is also seen in the US. Failure to adjust for educational level may result in residual confounding, especially if fractures are self-reported. Were you able to calculate total energy intake? What would be the limitations if no adjustment for total energy intake can be made? It is likely that other factors than age, oral glucocorticoids, hormone replacement therapy and previous fractures may confound the association between milk intake and fracture risk. Please provide arguments for not including other variables or update your analyses accordingly. 5. Information on dietary data was time-updated using repeat questionnaires, which is a strength of the study. Were covariates treated as time-updated covariates as well (you state “risk factors” are these the confounders)? Make sure to state in the tables what analyses are time-updated and which are not (milk vs total dairy). Also, did you perform sensitivity analyses using milk intake at baseline only (not time-updated analysis)? If so, what were the results? 6. Why did you use different reference categories for associations with fracture and associations with inflammation and bone turnover markers? Was the reference category selected before analysis? 7. The problem with the Kaplan-Meier curves presented as separate curves for the exposures is that these are not adjusted for confounders and might in this setting contribute to confusion. Please provide adjusted survival curves or consider omitting these stratified Kaplan-Meier curves. 8. The description of the population is somewhat unclear. How old were the recruited women? How many of the women ≥50 years had information on the exposure? How did you handle

	missing data? Please consider using a flow-chart as per STROBE guidelines. 9. It is not clear from the description how information on incident fractures and deaths were collected. Please specify. If fracture information was collected through self-report this should be clearly specified throughout the manuscript (including in the abstract) as it might have important implications for potential loss to follow-up. The problem with the self-reported fractures is usually not if those fractures identified are correctly reported (radiographically confirmed) but rather that you are likely to introduce a selection bias if you do not capture all incident fractures. Please also add a discussion regarding any advantages and disadvantages of the ascertainment method used. 10. It would be interesting to see the distribution of the amount of different dairy products (and other characteristics of the population) by total dairy intake in an additional table (corresponding to table 1 but across categories of dairy intake). This is important to understand what sources of dairy products are included in the different categories. 11. The discussion includes a large section on methodological discussions regarding A1 beta casein and galactose, with special reference to our Swedish study. The discussion on this topic is quite long considering your own conclusion that there is no association between milk intake and fracture risk. A1 beta casein is produced by some Swedish cow breeds, but not by all, and the dairy industry will not treat milk from different breeds differently. The source for fresh and fermented milk is the same and there seem to be no information that the protein content is different in milk and soured milk, which traditionally has been the major fermented milk product eaten in Sweden. Thus, content of A1 beta casein would not explain why milk intake is associated with a higher risk of hip fracture and soured milk/yoghurt intake is associated with a lower risk of hip fracture in our Swedish cohort. The evidence for A1 beta casein as being important for health outcomes seems to be scarce (https://www.ncbi.nlm.nih.gov/pubmed/30722004). 12. When it comes to the galactose discussion, there seems to be a misunderstanding regarding the galactose amount in different dairy products. In intestine, as well as in the fermentation process when producing soured milk and yoghurt, lactose is degraded by lactase, producing galactose and glucose. The concentration of free galactose is therefore higher in for instance yoghurt than in fresh milk. The total galactose load (the sum of galactose in lactose + free galactose) is however LOWER in fermented milk products compared to fresh milk. The Alm reference shows that this is furthermore dependent on storage time. The Richmond paper is not relevant in this context. The Abrahamson reference is a master's thesis (not peer reviewed; should be removed) and we have, together with Abrahamson and Abrahamson's thesis supervisor, published a short communication (peer reviewed; Ohlsson et al Int Dairy J 2017: http://dx.doi.org/10.1016/j.idairyj.2017.06.004) based on the results in the master's thesis showing that the total

	galactose load is lower in soured milk and yoghurt, compared with fresh milk. How much lower total galactose there is may be dependent on the starter culture used in fermentation, on live bacteria in the products and, as Alm showed, on storage time. Thus, the text regarding galactose content needs to be updated and clarified. 13. Bullet points: Strengths and limitations of the study: All epidemiological studies suffer from different kinds of biases to different degrees. The reader is not helped by the statement "The likelihood of bias is minimal due to random sample selection from the general population." Furthermore, generalizability is often regarded as a secondary issue when studying associations and we would first want to make sure there is internal validity before extrapolating results to other populations (you also discuss different sources of bias in the next items, suggest to remove this statement). In the discussion, be specific to what the results cannot be generalized to. Are they not generalizable to other women? To other women in Australia? Why is it then important that you have a random sample selection from the general population? 14. Discussion, page 13: "Acquiring the daily recommend calcium through diet is considered the easiest and safest lifestyle modification that could be achieved as a part of prevention and management of osteoporosis." Do you have a reference to support this statement? 15. Abstract and tables: Please state what fractures are included in MOF.

REVIEWER	Kristin Holvik Norwegian Institute of Public Health, Norway
REVIEW RETURNED	26-Jun-2019

GENERAL COMMENTS	General assessment: This manuscript reports the results from a cohort of Australian women, investigating the association between consumption of cow's milk and total dairy products, respectively, and risk of a major osteoporotic fracture defined as fracture of the hip, forearm, spine or proximal humerus. The Geelong Osteoporosis Study is a population-based study in South-Eastern Australia designed to investigate the epidemiology of osteoporosis. The study sample includes 833 women who were aged 50 years and older at baseline in 1993. Average follow-up was approximately 14 years and 206 incident fractures were identified. Compared with moderate milk-drinkers (<250 mg/day), they observed a trend towards an increased risk of a major osteoporotic fracture in non-drinkers of milk (HR 1.56, 95% CI 0.99, 2.46; Table 2) while there was a non-significant relative risk of 1.15 in the highest category of milk consumption (>500 mL/day). For dairy consumption, the lowest fracture risk was observed in the low-consumers, and there was a borderline 70% increased fracture risk in those consuming at least 800 grams per day compared with the reference group consuming 200-399 grams per day (Table 3). The cohort has been thoroughly described earlier (Pasco et al., Int J Epidemiol 2012;41(6):1565-75). The cohort includes both men and women, and it is not clear why women only have been included in the current study. Strengths of the study include the long follow-up, exposure ascertainment through food frequency questionnaires, the

	ability to discern non-milk-drinkers from those drinking low quantities, and that exposure data (milk consumption) and some covariates were available from several follow-up questionnaires and could be time-updated. The low sample size is a clear limitation. A priori power calculations were not performed, but the authors state that according to post hoc considerations, the study was powered to detect relative risks in the magnitude of 1.5 and higher, which is a large effect size for a dietary exposure. The manuscript concludes that “increased milk consumption is not associated with increased risk for major osteoporotic fracture”. With a sample size of 833 the study is severely underpowered to detect a true relative risk of 1.15, which is the point estimate reported for the highest milk intake category compared with the reference category in the main analysis. Specific points that need to be addressed: 1. More information is needed about the flow of participants. A total of 1494 women participated at baseline, but only 833 women aged 50 years and older were included in the current analyses. On pages 5-6 the number of participants at the respective follow-up cycles are presented. These are already published in the cohort profile paper (Pasco et al., Int J Epidemiol 2012;41(6):1565-75), and as far as I understand these numbers apply to the total cohort of women aged 20 years and older. For the current study it is of interest to present the numbers and percentages of women in the age range 50+ who were selected, invited, participated/rejected, and dropped out. Preferably, the attendance rate for the pertinent age range should be given for each follow-up cycle. Were there no exclusions based on e.g. missing dietary information? A STROBE checklist has been provided, but point #13a concerning reporting of participants has been answered with N/A while points #13b and #13c concerning reasons for non-participation and flow diagram have been skipped. It would be helpful to include a flow diagram to visualize the number of eligible participants and the number of events, deaths, emigrations and dropouts for each follow-up cycle in the pertinent age range. Also note that as of now, the apparent logic of the presentation (lines 1-3 on page 6) may give the confusing impression that the study sample was selected on the basis of completing the 10-year follow-up, which I assume is not the case. 2. I also wonder about the reason for restricting the current study to women, when a corresponding number of men were recruited to the cohort. 3. Given the overall evidence on the topic, why did the authors hypothesize an increased risk for major osteoporotic fracture at higher milk and dairy consumption? 4. Description of the fracture ascertainment needs some more detail. The manuscript states that “radiological reports were used to identify and confirm post-baseline incident fractures using a method that has been validated for use in the study region”. The formulation “identify and confirm” is ambiguous. Were fractures occurring since the previous follow-up self-reported in questionnaires at each follow-up cycle, and subsequently individually confirmed (or disproved) in radiologic reports? Were all individuals in the cohort linked to data from major radiology centres through personal identification numbers, ensuring that any incident fracture in any participant who had received an X-ray was captured? Or were these methods

	combined? For self-reported fractures, the concern is that fractures that happened between follow-up cycles but were not reported due to incomplete questionnaires, deaths, emigration/moving or dropout of various reasons, would be misclassified as no fracture. One may expect deterioration in health and mobility after a fracture injury to increase the risk of dropout. Please provide some brief details about the validity of the method and the expected degree of misclassification. 5. The term “total dairy consumption” implies a strong assumption about complete coverage of individual dairy intake. I am not familiar with dietary habits among middle-aged and older women in South-Eastern Australia, but since all participants have completed FFQ, the relative contribution of various sources to dairy consumption may be scrutinized. Could the authors suspect other dietary sources to contribute substantially and merit inclusion in the variable (milk content of e.g. porridge, pancakes, etc.?), or justify that the sources included represent total dairy intake? 6. The study question and analyses concerning total dairy consumption and milk consumption appear equally important in the manuscript. For transparency concerning the size of the groups and potential confounding, it would be helpful to include a background table showing the distribution of participants’ characteristics across categories of total dairy consumption. 7. In the statistical analyses, fracture risk in milk consumption categories were compared with the second lowest milk consumption category (<250 mL/day). The number of participants is highest in the reference category (n=393; 47% of the sample). The high-milk-consumers (>500 mL/day) represented 10% of the participants (n=84) and provided 14% of incident fractures (n=29). The lack of statistical power is already a cause of concern and the analyses comparing exposure categories do not utilize the potential statistical power in the entire sample. Was any attempt made to investigate risk across the continuous distribution of milk intake? To my understanding, the categories of milk used in the statistical analyses corresponded to the predefined categories in the questionnaire (only slightly modified). Total dairy intake, however, was calculated in grams per day from reported intakes of milk, cheese, yogurt and ice cream. Please clarify why calculated total dairy intake was treated as a categorical exposure variable with four categories, and whether it was attempted to investigate fracture risk across the continuous distribution of dairy consumption. Also, the choice of cut-offs and choice of reference category need to be justified. Was the second lowest category as reference predefined for both the milk and total dairy exposure, or was it based on group size or observed fracture rates in categories? 8. Statistical analysis: Please provide more technical details concerning how the time updated exposure and covariates (baseline, 6y and 10y follow-up) were handled in Cox regression. 9. Table 3: Fracture rates presented in the lowest, second highest and highest consumption category are incorrect. 10. The time points in question for the different analyses should be unambiguous. In the statistical analysis section, abstract, and results, it should be made clear that the linear regression analyses of CRP and bone turnover markers on milk consumption was based

	on cross-sectional data at baseline. Referring to these regression data as “the effects” of milk and dairy on bone turnover markers should be avoided. 11. The manuscript displays a somewhat exaggerated confidence in the representativeness and validity of the data. The sample size is consequently described as large. E.g.: “The likelihood of bias is minimal” (pages 3 and 16); “large representative sample of Australian women” (page 5). These statements should be modified, and the authors should take care discussing the limitations of the sample. 12. Minor suggestions and spelling errors: a) Major osteoporotic fractures should be defined in the abstract b) Page 4: inversly, should be inversely c) Page 4: The correct publication year of ref.18 is 2011, not 2010 d) Page 7: the term “interrogated” could be replaced by e.g., asked or enquired e) Page 11, 13, 16, 29 (footnote 1): PINP, should be P1NP

VERSION 1 – AUTHOR RESPONSE

Reviewer# 1
Points that need to be addressed:
1. More information is needed about the flow of participants. A total of 1494 women participated at baseline, but only 833 women aged 50 years and older were included in the current analyses. On pages 5-6 the number of participants at the respective follow-up cycles are presented. These are already published in the cohort profile paper (Pasco et al., Int J Epidemiol 2012;41(6):1565-75), and as far as I understand these numbers apply to the total cohort of women aged 20 years and older. For the current study it is of interest to present the numbers and percentages of women in the age range 50+ who were selected, invited, participated/rejected, and dropped out. Preferably, the attendance rate for the pertinent age range should be given for each follow-up cycle. Were there no exclusions based on e.g. missing dietary information? A STROBE checklist has been provided, but point #13a concerning reporting of participants has been answered with N/A while points #13b and #13c concerning reasons for non-participation and flow diagram have been skipped. It would be helpful to include a flow diagram to visualize the number of eligible participants and the number of events, deaths, emigrations and dropouts for each follow-up cycle in the pertinent age range. Also note that as of now, the apparent logic of the presentation (lines 1- 3 on page 6) may give the confusing impression that the study sample was selected on the basis of completing the 10-year follow-up, which I assume is not the case.
Response: Information on how participants were selected for the analysis and how missing data has been handled is included in the methods section. Besides, text in the manuscript has been amended accordingly. A participant flow diagram (Figure 1) has been constructed and attached to the manuscript. This includes information on the number of participants in each follow-up wave, participants had information on exposure data and participant who were lost to follow-up due to leaving the region. Additionally, an updated STROBE checklist has been attached.
2. I also wonder about the reason for restricting the current study to women, when a corresponding number of men were recruited to the cohort.
Response: We didn’t complete this same analysis in men due to insufficient power. Men are less likely to sustain a fracture, and we have a shorter follow-up time for men (from 2001 at the earliest). This would not provide a sufficient number of fractures to conduct a meaningful analysis. However, we might perform an analysis in a few years when the follow-up period has extended and there are more fracture outcomes.

3. Given the overall evidence on the topic, why did the authors hypothesize an increased risk for major osteoporotic fracture at higher milk and dairy consumption?
Response: The overall evidence on the topic is still from observational studies and meta-analysis, however, not from clinical studies. Besides, it is unclear to some extent whether the results from these observational studies (Swedish mammography cohort, or US cohorts) can be extrapolated to Australian women. Except the Swedish cohort study, which found a positive association between increased milk intakes and fractures, other studies that found no association between increased milk consumption and fractures have not been supported by a rigorous mechanism.
Considering, that milk/dairy contains substantial quantities of D-galactose and A1 beta-casein protein, and cognisant that these components are deleterious for health (from pre-clinical and epidemiological evidence), it provided the rationale to hypothesise that increasing the milk/dairy might be deleterious bone heath.
4. Description of the fracture ascertainment needs some more detail. The manuscript states that "radiological reports were used to identify and confirm post-baseline incident fractures using a method that has been validated for use in the study region". The formulation "identify and confirm" is ambiguous. Were fractures occurring since the previous follow-up self-reported in questionnaires at each follow-up cycle, and subsequently individually confirmed (or disproved) in radiologic reports? Were all individuals in the cohort linked to data from major radiology centres through personal identification numbers, ensuring that any incident fracture in any participant who had received an X-ray was captured? Or were these methods combined? For self-reported fractures, the concern is that fractures that happened between follow-up cycles but were not reported due to incomplete questionnaires, deaths, emigration/moving or dropout of various reasons, would be misclassified as no fracture. One may expect deterioration in health and mobility after a fracture injury to increase the risk of dropout. Please provide some brief details about the validity of the method and the expected degree of misclassification.
Response: Fractures were not self-reported. Further information on how fractures were ascertained has now been included in the methods section.
"Post-baseline incident fractures were identified using a method that have been validated for fracture ascertainment in the region. Radiological reports (X-ray) of fractures from all radiological centres in the region were scrutinised to identify and confirm fractures. Trained research personnel examined each record individually and determined the most appropriate international code of diseases version 9 (ICD-9) codes for fracture site, as well as level of trauma. Codes were not directly extracted from medical records"
Additionally, the advantages and disadvantages with this method of fracture ascertainment has been included in the discussions section

5. The term "total dairy consumption" implies a strong assumption about complete coverage of individual dairy intake. I am not familiar with dietary habits among middle-aged and older women in South-Eastern Australia, but since all participants have completed FFQ, the relative contribution of various sources to dairy consumption may be scrutinized. Could the authors suspect other dietary sources to contribute substantially and merit inclusion in the variable (milk content of e.g. porridge, pancakes, etc.?), or justify that the sources included represent total dairy intake?
Response: When defining participants' total dairy consumption for this study, we included the key components in dairy e.g. milk, all forms of cheese and yogurt, and ice-cream. Accounting for milk, cheese, yogurt and ice-cream would represent the predominant forms of dairy consumed in Australia, however, will not be a precise measure for ones' total dairy intake.
The questions in the dietary questionnaire are framed to capture the habitual dietary intake of people. Therefore, when participants are asked the question "what type of milk do you usually use" and follows with the question "how much milk do you usually consume each day" this encourages participants to report the approximate quantity of milk they consume on a daily basis, which may also

include the usage of milk in cereals, tea and coffee, cooking, etc in the form of cups consumed per day. This explanation also applies for other dairy products (cheese, yogurt and ice-cream) that we have included for accounting total dairy. Therefore, based on the dietary questions and the components accounted for capturing total dairy intake- this would represent rough estimates for total dairy consumption.

6. The study question and analyses concerning total dairy consumption and milk consumption appear equally important in the manuscript. For transparency concerning the size of the groups and potential confounding, it would be helpful to include a background table showing the distribution of participants' characteristics across categories of total dairy consumption.

Response: A descriptive table (supplementary Table 1) has been created for participants characteristics based on total dairy consumption groups.

7. In the statistical analyses, fracture risk in milk consumption categories were compared with the second lowest milk consumption category (<250 mL/day). The number of participants is highest in the reference category (n=393; 47% of the sample). The high-milk-consumers (>500 mL/day) represented 10% of the participants (n=84) and provided 14% of incident fractures (n=29). The lack of statistical power is already a cause of concern and the analyses comparing exposure categories do not utilize the potential statistical power in the entire sample. Was any attempt made to investigate risk across the continuous distribution of milk intake?

To my understanding, the categories of milk used in the statistical analyses corresponded to the predefined categories in the questionnaire (only slightly modified). Total dairy intake, however, was calculated in grams per day from reported intakes of milk, cheese, yogurt and ice cream. Please clarify why calculated total dairy intake was treated as a categorical exposure variable with four categories, and whether it was attempted to investigate fracture risk across the continuous distribution of dairy consumption. Also, the choice of cut-offs and choice of reference category need to be justified. Was the second lowest category as reference predefined for both the milk and total dairy exposure, or was it based on group size or observed fracture rates in categories?

Response: We did not investigate risk across the continuous distribution of milk intake as the data on milk consumption were captured by pre-determined categories and this precluded a meaningful assessment with milk as a continuous variable. In terms of total dairy, the distribution/nature of the data (clustered) was not suitable to proceed with total dairy as a continuous variable. Also, when total dairy categorise were generated, the distribution of participant numbers in each category was taken into account. The second lowest category as reference was chosen for both the milk and total dairy exposure based on considering the group size.

8. Statistical analysis: Please provide more technical details concerning how the time updated exposure and covariates (baseline, 6y and 10y follow-up) were handled in Cox regression.

Response: Information on how confounders were time updated in follow-up waves have been included in the manuscript under the statistical analysis section.

"The final model consisted, age, oral glucocorticoid use, HT use and pre-baseline factures as confounders. Information on milk consumption, oral glucocorticoid use, and HT use were time updated at the 6 and 10-year follow-up. Age was time updated in all follow-up waves. Information on pre-baseline fractures were not time updated and kept constant for the analysis"

9. Table 3: Fracture rates presented in the lowest, second highest and highest consumption category are incorrect.

Response: The fracture rates that were incorrect have now been corrected.

10. The time points in question for the different analyses should be unambiguous. In the statistical analysis section, abstract, and results, it should be made clear that the linear regression analyses of CRP and bone turnover markers on milk consumption was based on cross-sectional data at baseline. Referring to these regression data as "the effects" of milk and dairy on bone turnover markers should be avoided.

Response: In response to this comment, the analysis between milk/total dairy, and CRP and bone turnover markers has been referred to as "cross-sectional at baseline" in the abstract, methods, results and discussion section. The heading of table 4 has been amended accordingly to reflect this

change and the phrase “the effects” has been removed.
11. The manuscript displays a somewhat exaggerated confidence in the representativeness and validity of the data. The sample size is consequently described as large. E.g.: “The likelihood of bias is minimal” (pages 3 and 16); “large representative sample of Australian women” (page 5). These statements should be modified, and the authors should take care discussing the limitations of the sample.
Response: The text in the manuscript has been amended accordingly and the potential limitations of the sample have now been discussed.
12. Minor suggestions and spelling errors:
1. a) Major osteoporotic fractures should be defined in the abstract
2. b) Page 4: inversly ◊ inversely
3. c) Page 4: The correct publication year of ref. 18 is 2011, not 2010
4. d) Page 7: the term “interrogated” could be replaced by e.g., asked or enquired
5. e) Page 11, 13, 16, 29 (footnote 1): PINP ◊ P1NP
Response: All these minor comments have been addressed in the manuscript.
Reviewer #2:
I have the following questions and comments to the authors:
1. Although osteoporosis is associated with a marked increased risk of fracture, most fractures are still preceded by a fall. The largest number of fractures also occur among those who do not have osteoporosis. The introduction should reflect this as well as the bone aspects of fracture risk.
Response: The introduction section has been edited to address this comment. Text has been added to describe that most fractures are preceded by falls and many are not the result of osteoporosis.
2. The meta-analysis cited in ref 18 is rather old and could be excluded. More importantly, results from meta analyses on dietary studies are complicated because of the different dietary context and the different range of intakes (see for instance Barnard ND, Willett WC, Ding EL. The Misuse of Meta-analysis in Nutrition Research. JAMA2017;318:1435-1436. https://doi.org/10.1001/jama.2017.12083). Reference 19 is problematic due to the large heterogeneity between the studies in the main analysis (largest intake group vs the smallest intake group) and the dose-response analysis shows a different result than what is referred to in the introduction. The heterogeneity between studies is due to the differences in the underlying studies with regards to range of milk intake, number of fracture cases, what confounders were adjusted for, and how fractures were ascertained. I encourage the authors to highlight the problems with meta analyses in this context.
Response: Reference 18 (the old meta-analysis) has been removed. The issues associated with the meta-analysis conducted by Bian et al have now been elaborated in the introduction section.
3. The interpretation of the results seems to rely heavily on statistical significance. In the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (by ICMJE) and in the STROBE guidelines it is said that authors should avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size. I think that considering the small number of fractures in the present study it would be more fair to describe your association between milk intake (and total dairy intake) and fracture risk as u- (or perhaps J-) shaped with higher risks in the no consumption and the high consumption groups and the lowest risk in the low-intermediate consumption groups, however certainty is limited because of the small number of cases resulting in low precision in the estimates. Further, P-values in table 1 is discouraged (as per STROBE guidelines).
Response: P values in Table 1, 2, 3, and 4 have been removed. Besides, we have used hazard ratios to interpret the results. We have also added an explanation as to why we did not describe the results as “j-shaped” in the Discussion section.
“We were unable to describe the association between milk/total dairy intake and fracture risk, as U/J shaped graphs showing higher risks in the zero and high consumption groups and the lowest risk in the low-intermediate consumption groups because of the low number of fractures, which may lead to lower precision in the estimates”
4. How did you select which confounders to adjust for? There seems to be an association between

milk intake and educational level – this is also seen in the US. Failure to adjust for educational level may result in residual confounding, especially if fractures are self-reported. Were you able to calculate total energy intake? What would be the limitations if no adjustment for total energy intake can be made? It is likely that other factors than age, oral glucocorticoids, hormone replacement therapy and previous fractures may confound the association between milk intake and fracture risk. Please provide arguments for not including other variables or update your analyses accordingly.
Response: Information on selecting confounders has been included in the statistical analysis section.
"Covariates (BMD, BMI, smoking, alcohol consumption, pre-baseline fractures incidents, diabetes, IRSD, education, mobility, medications that influence bone metabolism, calcium and vitamin D supplements) were assessed in bivariate Cox regression analysis to determine their impacts on the association between milk/total dairy consumption and fractures. The covariates that impacted the hazard ratio when added or removed (considering the statistical significance and change of HR in the exposure of interest) from the model were included in the final Cox regression model. In addition, when deciding on the confounders, the potential of the covariate to be associated with both the exposure and outcome was also considered. The final model consisted, age, oral glucocorticoid use, HT use and pre-baseline factures as confounders"
We have assessed the potential of educational level as a potential confounder; however, its impact was negligible as a confounder. Therefore, educational level was not adjusted in the final Cox regression model. The text in the document has been amended to highlight this point.
We did not retain information on total energy intake; however, we did have information on BMI and physical activity level, which we considered as surrogate measures for total energy intake and were assessed in the bivariate Cox models. However, these measures did not qualify as confounders and wasn't included in the final Cox regression model.
5. Information on dietary data was time-updated using repeat questionnaires, which is a strength of the study. Were covariates treated as time-updated covariates as well (you state "risk factors" are these the confounders)? Make sure to state in the tables what analyses are time-updated and which are not (milk vs total dairy). Also, did you perform sensitivity analyses using milk intake at baseline only (not time-updated analysis)? If so, what were the results?
Response: Additional information regarding, which covariates were time updated has been included in the statistical analysis section of the manuscript. Additionally, information on time updated confounders were added as footnotes to Table 2 and Table 3 to enhance clarity.

"Information on milk consumption, oral glucocorticoid use, and HT use were time updated at the 6 and 10-year follow-up. Age was time updated in all follow-up waves. Information on pre-baseline fractures were not time updated and kept constant for the analysis. We also performed a multivariable adjusted sensitivity analysis using baseline milk values only"
We did perform a sensitivity analysis and the following text has been included in the manuscript as response to this comment.
"The multivariable adjusted sensitivity analysis, which was performed using baseline milk values only resulted non-significant higher HR ratio for non- milk consumers (HR:1.53; CI: 0.96-2.44; p=0.07) and >500 mL/d of milk consumers (HR:1.13; CI:0.74-1.72; p=0.58) compared to consuming < 250 mL/d milk of milk"
6. Why did you use different reference categories for associations with fracture and associations with inflammation and bone turnover markers? Was the reference category selected before analysis?
Response: We used <250 mL/d, < 200-399 g/d as reference categories in the Cox-regression, considering the group size in each category (these categories had the highest observations in terms of exposure). However, when performing the cross-sectional regression at baseline, we used the lower end of the milk/total dairy categories (no milk, < 200 g/d) as the reference categories as we were interested in assessing the associations between milk/total dairy and serum markers (inflammatory and bone turnover) in higher end categories, compared to the lowest categories
7. The problem with the Kaplan-Meier curves presented as separate curves for the exposures is that these are not adjusted for confounders and might in this setting contribute to confusion. Please

provide adjusted survival curves or consider omitting these stratified Kaplan-Meier curves.
Response: We acknowledge the fact that Kaplan-Meier curves were not adjusted for confounders, however we chose to keep them in the manuscript as a useful descriptive tool to illustrate incidence rates over study time. We specifically preferred this to model-adjusted survival curves so that readers could have a comprehensive description of the crude outcome progress to compare with finding from the model in order to compare the model-based findings (which are based on some statistical assumptions) with the actual data. We added a footnote for KM plot to explain this.
8. The description of the population is somewhat unclear. How old were the recruited women? How many of the women ≥50 years had information on the exposure? How did you handle missing data? Please consider using a flow-chart as per STROBE guidelines.
Response: A flow chart (Figure 1) has now been included in the manuscript to show the numbers of participants in the study at different time points. Additional text has also been included in the methods section to show how missing data was handled.
"For the purposes of the analysis women only ≥50yr at baseline were considered. Of the 836 women aged ≥ 50 yr, 833 women were included in the analysis after excluding records with missing information on milk intake (Figure1)"
9. It is not clear from the description how information on incident fractures and deaths were collected. Please specify. If fracture information was collected through self-report this should be clearly specified throughout the manuscript (including in the abstract) as it might have important implications for potential loss to follow-up. The problem with the self-reported fractures is usually not if those fractures identified are correctly reported (radiographically confirmed) but rather that you are likely to introduce a selection bias if you do not capture all incident fractures. Please also add a discussion regarding any advantages and disadvantages of the ascertainment method used.
Response: The incident fractures were not self-reported. We have added additional information in the methods section to explain how fractures and deaths have been ascertained. Additionally, we have inserted text in the discussion section to indicate the advantages and disadvantages of the fracture ascertainment method.
10.It would be interesting to see the distribution of the amount of different dairy products (and other characteristics of the population) by total dairy intake in an additional table (corresponding to table 1 but across categories of dairy intake). This is important to understand what sources of dairy products are included in the different categories.
Response: Participant characteristics have been described based on the total dairy consumption categories. We have provided this table separately (Supplementary Table 1) and would prefer to have this as a supplementary material.
11.The discussion includes a large section on methodological discussions regarding A1 beta casein and galactose, with special reference to our Swedish study. The discussion on this topic is quite long considering your own conclusion that there is no association between milk intake and fracture risk. A1 beta casein is produced by some Swedish cow breeds, but not by all, and the dairy industry will not treat milk from different breeds differently. The source for fresh and fermented milk is the same and there seem to be no information that the protein content is different in milk and soured milk, which traditionally has been the major fermented milk product eaten in Sweden. Thus, content of A1 beta casein would not explain why milk intake is associated with a higher risk of hip fracture and soured milk/yoghurt intake is associated with a lower risk of hip fracture in our Swedish cohort. The evidence for A1 beta casein as being important for health outcomes seems to be scarce (https://www.ncbi.nlm.nih.gov/pubmed/30722004).
Response: The discussion section has been amended to reduce the text regarding A1 beta-casein.
12.When it comes to the galactose discussion, there seems to be a misunderstanding regarding the galactose amount in different dairy products. In intestine, as well as in the fermentation process when producing soured milk and yoghurt, lactose is degraded by lactase, producing galactose and glucose. The concentration of free galactose is therefore higher in for instance yoghurt than in fresh milk. The total galactose load (the sum of galactose in lactose + free galactose) is however LOWER in fermented milk products compared to fresh milk. The Alm reference shows that this is furthermore

dependent on storage time. The Richmond paper is not relevant in this context. The Abrahamson reference is a master's thesis (not peer reviewed; should be removed) and we have, together with Abrahamson and Abrahamson's thesis supervisor, published a short communication (peer reviewed; Ohlsson et al Int Dairy J 2017: http://dx.doi.org/10.1016/j.idairyj.2017.06.004) based on the results in the master's thesis showing that the total galactose load is lower in soured milk and yoghurt, compared with fresh milk. How much lower total galactose there is may be dependent on the starter culture used in fermentation, on live bacteria in the products and, as Alm showed, on storage time. Thus, the text regarding galactose content needs to be updated and clarified.
Response: We are grateful to the reviewer for providing more insightful information on this topic, which clearly describes the total galactose load in fermented dairy. Text in the discussion section in the manuscript has been edited in response to these comments.
13.Bullet points: Strengths and limitations of the study: All epidemiological studies suffer from different kinds of biases to different degrees. The reader is not helped by the statement "The likelihood of bias is minimal due to random sample selection from the general population." Furthermore, generalizability is often regarded as a secondary issue when studying associations and we would first want to make sure there is internal validity before extrapolating results to other populations (you also discuss different sources of bias in the next items, suggest to remove this statement). In the discussion, be specific to what the results cannot be generalized to. Are they not generalizable to other women? To other women in Australia? Why is it then important that you have a random sample selection from the general population?
Response: The text in the discussion section has been edited and updated in response to the comment made (see below). Also, the discussion on other different sources of bias has been removed.
"Our study has several strengths. One strength of the study is, GOS comprise a randomly selected group of participants, which has shown to be similar to the broader Australian population (in terms of income, SES, etc.). However, our findings are not generalisable to men, nor other countries with different distributions of ethnicities, diet patterns and many other factors. However, our study results are likely generalisable to the broader Australian population of women"
14.Discussion, page 13: "Acquiring the daily recommend calcium through diet is considered the easiest and safest lifestyle modification that could be achieved as a part of prevention and management of osteoporosis." Do you have a reference to support this statement?
Response: A reference has been added to the manuscript to support this statement.
15.Abstract and tables: Please state what fractures are included in MOF.
Response: The fractures included in the category "MOF" (major osteoporotic fractures) have been listed in the abstract and tables.

VERSION 2 – REVIEW

REVIEWER	Liisa Byberg Uppsala University, Sweden
REVIEW RETURNED	24-Sep-2019

GENERAL COMMENTS	The authors have provided satisfactory responses and amendments to their manuscript. Thank you. I have only a couple of minor remarks. In table 2, the estimates for no (vs reference) milk consumption, both the age-adjusted (**1.54 (0.98, 2.44)) and the multivariable adjusted (**1.56 (0.99, 2.46)) estimates are marked with ** indicating p<0.05. However the 95% confidence intervals cover 1 and hence p would be >0.05? Also, in the results text, the p-values are said to be 0.06. Could the authors please double check that what numbers are correct?

	On page 17 "… could not differentiate between conventional dairy products or A2 milk products; thus, we were not able to investigate particular milk proteins as mediating potential negative effects associated with milk" The general reader will probably not know what an A2 product is, needs clarification (probably possible by just rephrasing the sentence somewhat) Throughout the text there are some minor type-o's; these are some that I found: page 13: values only resulted non-significant higher page 15: may augmented page 16: was not time update

REVIEWER	Kristin Holvik Norwegian Institute of Public Health
REVIEW RETURNED	09-Sep-2019

GENERAL COMMENTS	The manuscript has been revised to accommodate many of the suggestions from the reviewers. I appreciate the Authors' responses and amendments following my questions to the previous version, which have clarified some of my concerns. I particularly appreciate the text additions to the manuscript which help elucidate several aspects that were unclear in the previous version. The quality of the manuscript has improved after revision. However, I have three remaining comments. • Regarding the response to my previous point 3, concerning rationale for hypothesising an increased fracture risk at higher milk and dairy intake. The Introduction chapter states: "We hypothesised that increased milk and total dairy consumption may be associated with increased risk for MOF". What I am asking for is a clearly stated rationale for expecting this direction of association. The manuscript would benefit from including a statement in the introduction chapter to explain and justify why the authors expected a positive association in the GOS cohort. After all, this association has previously been observed for sweet milk and hip fracture in women in one observational study only. The same cohort exhibited an opposite direction of association, i.e. an apparent protective association, of yogurt and soured milk for hip fractures. Based on those results, I do not find it evident that one would expect an increased fracture risk for higher total dairy consumption. • Regarding the response to my previous point 5, concerning dairy consumption: If this is correct, it is informative to know that the milk variable was explicitly not restricted to drinking milk but also covered milk added to coffee and other foods and dishes, and that the same goes for the cheese and yogurt questions. In most FFQs, consumption of cooked meals that may have a high content of cheese (pizza, soufflés, mac-and-cheese, omelets etc.) would be reported separately and not captured by a cheese question standing alone. Correspondingly, composed meals with a high milk content would generally be reported separately from drinking milk. Does the FFQ completed by the GOS participants differ in these aspects? This is not clear from the manuscript and it would be nice to specify this information.

	• Regarding the response to my previous point 7, concerning the choice of performing statistical analysis across categories of consumption rather than across the continuous distribution. I recommend explaining in the Methods section that: a) Analysis with categories of total dairy was chosen due to the clustered nature of the data making it unfeasible to treat the exposure as a continuous variable, and b) The second lowest category as reference was chosen for both the milk and total dairy exposure based the number of participants within categories.

VERSION 2 – AUTHOR RESPONSE

Reviewer:2

1. Regarding the response to my previous point 3, concerning rationale for hypothesising an increased fracture risk at higher milk and dairy intake. The Introduction chapter states: "We hypothesised that increased milk and total dairy consumption may be associated with increased risk for MOF". What I am asking for is a clearly stated rationale for expecting this direction of association. The manuscript would benefit from including a statement in the introduction chapter to explain and justify why the authors expected a positive association in the GOS cohort. After all, this association has previously been observed for sweet milk and hip fracture in women in one observational study only. The same cohort exhibited an opposite direction of association, i.e. an apparent protective association, of yogurt and soured milk for hip fractures. Based on those results, I do not find it evident that one would expect an increased fracture risk for higher total dairy consumption.

Response: We apologise for not making this point clear during the last round of revision. In response to the above comment the text has been amended and included in the Introduction with relevant references.
Also, we have explicitly stipulated the reason for hypothesizing why increased milk consumption is associated with increased risk for MOF and removed total dairy from this sentence.
Following this, in a separate paragraph, we have clarified the reason for assessing the association between total dairy consumption and MOF.

"Although the overall evidence on increased milk intake appears supportive of reducing fractures, dissecting milk further to the molecular level demonstrates that milk contains compounds such as D-galactose (milk sugar) and A1-beta-casein (mutated protein variant) that may be detrimental to bone health (26-28). Pre-clinical studies show that these compounds are implicated in inflammation and oxidative stress pathways that can negatively impact bone metabolism (29, 30). Moreover, Pasco et al. previously indicated that increased milk intake is associated with depressive disorder (31), a condition that is comorbid with fractures (32, 33). Therefore, we hypothesised that increased milk consumption may be associated with increased risk for MOF by triggering inflammation and oxidative stress.
Other milk derived products such as yogurt and cheese have a distinct biological profile to milk and may have a protective role in bone health due to the presence of probiotics, prebiotics and other bioactive compounds; these in turn have the potential to attenuate inflammation and oxidative stress (34, 35). Studies have assessed the effects of these products on bone separately to milk; however, the synergistic impact of dairy products (including milk, yogurt, cheese, ice-cream) with different molecular and biological profiles is poorly unravelled. Therefore, we aimed to assess the association between total dairy consumption and MOF in women"

2. Regarding the response to my previous point 5, concerning dairy consumption: If this is correct, it is

informative to know that the milk variable was explicitly not restricted to drinking milk but also covered milk added to coffee and other foods and dishes, and that the same goes for the cheese and yogurt questions. In most FFQs, consumption of cooked meals that may have a high content of cheese (pizza, soufflés, mac-and-cheese, omelets etc.) would be reported separately and not captured by a cheese question standing alone. Correspondingly, composed meals with a high milk content would generally be reported separately from drinking milk. Does the FFQ completed by the GOS participants differ in these aspects? This is not clear from the manuscript and it would be nice to specify this information.

Response: We are grateful for the opportunity to further clarify this point. The questionnaire used to capture information related to dairy was not exclusively designed to capture information on drinking milk or cheese/yogurt consumed on its own. The questions are, designed to include milk, cheese, and yogurt consumed in all forms e.g milk in pasta/coffee, cheese in pizza. Due to the questionnaire being self-administered, there might be errors associated in reporting and this is a well-recognised limitation of nutritional epidemiology. For instance, when the question "how much of milk you consume each day" is asked, one may not recall milk added for baking pasta or cake. Therefore, this might lead to some under reporting and will be a potential limitation.

The following phrases are included in the Method section to elaborate the intention of the questionnaire, which was used to capture overall milk, yogurt and cheese intake on a daily basis (page 7).

"Participants were asked questions about the habitual/type of (all forms e.g. milk used in cooking, baking and in coffee) milk consumed (whole, reduced fat, calcium fortified, soy, goat's milk, butter milk, and evaporated) and the quantity consumed each day"

"Information on other dairy products such as cheese, yogurt (all forms e.g. cheese, and yogurt used in cooking, baking)…"

Additionally, we have included examples that were used to differentiate hard (cheddar and tasty) and soft (cream, cottage) cheese consumption in the questionnaire and this is included in the Methods section of the manuscript (page 7 & 8).

Finally, the following sentences have been incorporated into the manuscript under the Discussion section to address the limitation.
"Although the dietary questionnaire was designed to provide information on participants' habitual dairy intake, it is possible that dairy contained in manufactured/prepared products is not captured and thereby it underestimates total dairy consumption"

3. Regarding the response to my previous point 7, concerning the choice of performing statistical analysis across categories of consumption rather than across the continuous distribution. I recommend explaining in the Methods section that: a) Analysis with categories of total dairy was chosen due to the clustered nature of the data making it unfeasible to treat the exposure as a continuous variable, and b) The second lowest category as reference was chosen for both the milk and total dairy exposure based the number of participants within categories.

Response: The following sentences has been added to the manuscript in the Methods section to address the above comment
"The clustered nature of total dairy distribution, made it unfeasible to consider it as a continuous variable for analytical purpose, and as such it was treated as categorical variable in the analysis and categorised as < 200 g/d, 200-399 g/d, 400-700 g/d, ≥800 g/d. The second lowest category was

chosen as reference for total dairy because it was the largest group": page 8
"The second lowest category was chosen as reference for milk consumption as this category benefits robustness due to higher number of participants within the category": page 7

Reviewer: 1

1. In table 2, the estimates for no (vs reference) milk consumption, both the age-adjusted (**1.54 (0.98, 2.44)) and the multivariable adjusted (**1.56 (0.99, 2.46)) estimates are marked with ** indicating p<0.05. However, the 95% confidence intervals cover 1 and hence p would be >0.05? Also, in the results text, the p-values are said to be 0.06. Could the authors please double check that what numbers are correct?

Response: we are very grateful for alerting us to this typo. The footnote for Table 1 mentioning "p< 0.05" is removed and we confirm the p values in Results is correct. Additionally, the typographical error in both Table 2 footnote and the relevant text in Methods has been corrected

2.On page 17
" could not differentiate between conventional dairy products or A2 milk products; thus, we were not able to investigate particular milk proteins as mediating potential negative effects associated with milk"
The general reader will probably not know what an A2 product is, needs clarification (probably possible by just rephrasing the sentence somewhat)

Response: The following sentence has been included in order to avoid confusion and text has been amended in this paragraph to improve the logical flow in this section (page 17 & 18)

"In addition, when querying about the type of milk consumed, A2 milk/milk products (which contains exclusively A2 milk proteins) were not provided as an option to be selected by the participant; thus, we were not able to investigate particular milk proteins as potential mediators in the association with milk consumption"

Throughout the text there are some minor type-o's; these are some that I found:
page 13: values only resulted non-significant higher
page 15: may augmented
page 16: was not time update

Response: These errors have been corrected, thank you

VERSION 3 – REVIEW

REVIEWER	Liisa Byberg Uppsala University, Sweden
REVIEW RETURNED	15-Oct-2019

GENERAL COMMENTS	I have no further comments on the manuscript.

REVIEWER	Kristin Holvik Norwegian Institute of Public Health, Norway
REVIEW RETURNED	16-Oct-2019

GENERAL COMMENTS	My queries have been answered and I have no further points.